

# On Choosing Committees Based on Approval Votes in the Presence of Outliers

Palash Dey, Neeldhara Misra, and Y. Narahari

Indian Institute of Science, Bangalore

`{palash,neeldhara,hari}@csa.iisc.ernet.in`

November 16, 2015

## Abstract

We study the computational complexity of committee selection problem for several approval-based voting rules in the presence of outliers. Our first result shows that outlier consideration makes committee selection problem intractable for approval, net approval, and minisum approval voting rules. We then study parameterized complexity of this problem with five natural parameters, namely the target score, the size of the committee (and its dual parameter, the number of candidates outside the committee), the number of outliers (and its dual parameter, the number of non-outliers). For net approval and minisum approval voting rules, we provide a dichotomous result, resolving the parameterized complexity of this problem for all subsets of five natural parameters considered (by showing either FPT or W[1]-hardness for all subsets of parameters). For the approval voting rule, we resolve the parameterized complexity of this problem for all subsets of parameters except one.

We also study approximation algorithms for this problem. We show that there does not exist any  $\alpha(\cdot)$  factor approximation algorithm for approval and net approval voting rules, for any computable function  $\alpha(\cdot)$ , unless  $P = NP$ . For the minisum voting rule, we provide a  $(1 + \varepsilon)$  factor approximation algorithm running in time  $n^{O(\log m/\varepsilon^2)}$ , for every constant  $\varepsilon > 0$ , where  $m$  and  $n$  are the number of candidates and the number of votes respectively.

## 1 Introduction

Aggregating preferences of agents is a fundamental problem in artificial intelligence and social choice theory [Conitzer, 2010]. Typically, agents (or voters) express their preferences over alternatives (or candidates). There are many different models for the expression of preferences, ranging from the simplistic (each voter provides his or her favorite choice, also known as *plurality voting*) to the comprehensive (each voter provides a complete ranking over the set of all candidates). *Approval ballots* are an intermediate model, where a voter approves or disapproves of each candidate — thus a vote may be thought of as a subset of approved candidates or as a binary string indexed by the candidate set. Approval votes are considered as a good compromise between the extreme models — they provide the agent an opportunity to make a comment about every candidate, without incurring the overhead of determining a full ranking on the candidate set [Fishburn, 1978, Brams, 2008, Kilgour, 2010, Lepelletier, 2013, Baumeister et al., 2010].

Our work in this paper focuses on finding the “best” subset of  $k$  candidates when given  $n$  approval votes over  $m$  candidates. The set of candidates together with all the votes is usually called an *election instance* or simply an election. We use the term *electorate* to refer to the set of votes of all the agents, and the term *committee* to refer to a subset of  $k$  candidates. Since a committee is a subset of candidates, and a vote can

also be interpreted as a subset of the candidate set<sup>1</sup>, one might consider various natural notions of distance between these two sets. A fixed notion of distance leads to a measure of suitability of a committee with respect to an election — for instance, by considering the sum of distances to all voters, or the maximum distance incurred from any voter. For a committee  $\mathcal{X}$  and a vote  $\mathcal{S}$ , the notions of distance that are well-studied in the literature include the size of the symmetric difference (leading to the *minisum* or *minimax* rules) [Brams et al., 2007, LeGrand et al., 2007, Caragiannis et al., 2010, Gramm et al., 2003], the size of the intersection (leading to the notion of *approval score*), or the difference between  $|\mathcal{X} \cap \mathcal{S}|$  and  $|\mathcal{X} \setminus \mathcal{S}|$  (leading to the notion of *net approval scores*) [Procaccia et al., 2008, Skowron and Faliszewski, 2015, Aziz et al., 2015].

**Motivation** The standard approach to the selection of a winning committee is to look for one that optimizes these scoring functions over the entire election — that is, when the sum or max of scores are taken over all the votes. However, it is plausible that for some elections, there is a committee that represents a very good consensus when the scores are taken over a *subset* of the voters rather than the entire set. For example, consider the approval voting rule mentioned above. The approval score of a committee  $\mathcal{X}$  is the sum of  $|\mathcal{X} \cap \mathcal{S}|$ , taken over all votes  $\mathcal{S}$  in the electorate. Let the *approval score per vote* be the ratio of the approval score of  $\mathcal{X}$  to  $n$ . For a committee of size  $k$ , the best approval score per vote possible is  $k$ . Clearly, the higher the approval score per vote, the greater a voter is ‘satisfied’ on average. Now, consider the following toy example. Let  $\mathcal{L} \subseteq [m]$  be an arbitrary but fixed subset of  $k$  candidates. Suppose the election has  $(n - 1)$  votes that are all exactly equal to  $\mathcal{L}$ , and one vote that approves  $[m] \setminus \mathcal{L}$ . The best approval score per vote that one can hope for here is  $k(1 - 1/n)$ , whereas if we restrict our attention to the election on the first  $(n - 1)$  votes alone, then  $\mathcal{L}$ ’s approval score per vote is  $k$ , which is now the best possible. Of course, the difference here is not significant for large  $n$ , but the illustration does provoke the following thought: is there a subset of at least  $n^*$  votes that admit a committee whose approval score per vote is at least  $k^*$ ? For  $n^* = n$ , we are back to the original question of finding the best committee. This more general setting, however, allows us to explore trade-offs and structure in the election: if there is a committee that prescribes a very good consensus over a large fraction of the election, then it is likely to be a more suitable choice for the community compared to a committee that optimizes the same score function over the entire election. While these committees may coincide (as in the toy example), there are easy examples where they would not, making this a question worth exploring.

The notion of finding a good consensus over a large subset has been explored in various contexts. In particular, the idea of using *outliers* is quite common in the literature of *Closest String* problems. The setting of closest string involves a collection of  $n$  strings, and the goal is to find a single string that minimizes either the maximum distance, or the sum of distances, from all the input strings. The most commonly studied notion of distance is the Hamming distance. Notice that once we interpret votes as binary strings, we are asking a very similar question, and the main distinction is that our search is only over strings that have a fixed number of ones. This similarity has been noted and explored in some works on voting before (see, for instance, [Byrka and Sornat, 2014]). In the context of Closest String, a question that is often asked is the following: given a budget  $k$ , is there a string that is “close to” at least  $k$  strings? This question has been studied for both the minimax and minisum notions of closeness [Boucher et al., 2013, Lo et al., 2014]. Typically, the strings that are left out are called *outliers*. In the context of voting, one imagines that there may be a few votes that express rather tangential opinions, and that a good consensus emerges once they are removed. We will also refer to such votes as outliers. In the context of social choice theory, the very notion of Young voting rule can be regarded as to finding minimum number of outliers whose removal makes some candidate Condorcet winner [Young and Levenglick, 1978, Bartholdi III et al., 1989a, Rothe et al., 2003, Betzler et al., 2010].

**Our Framework** One of the advantages of using scoring rules for approval ballots described above is that the winning committees are polynomially computable for most of them (with the notable exception

---

<sup>1</sup>This is simply the subset of approved candidates.

of the minimax voting rule). However, once we pose the question of whether a target score  $t$  is achievable after the removal of at most  $k$  outliers, the complexity landscape changes dramatically. We show that this question is a computationally hard problem — in particular, we establish NP-hardness. Having shown hardness in the classical setting, we explore the complexity further, primarily from the perspective of exact exponential algorithms and approximation algorithms. In the context of the former, we use the framework of parameterized complexity. Briefly, a parameterized problem instance comprises of an instance  $x$  in the usual sense, and a parameter  $k$ . A problem with parameter  $k$  is called *fixed parameter tractable* (FPT) if it is solvable in time  $f(k) \cdot p(|x|)$ , where  $f$  is an arbitrary function of  $k$  and  $p$  is a polynomial in the input size  $|x|$ . While there have been important examples of traditional algorithms that have been analyzed in this fashion, the theoretical foundations for deliberate design of such algorithms, and a complementary complexity-theoretic framework for hardness, were developed in the late nineties [Downey and Fellows, 1988, Downey and Fellows, 1995a, Abrahamson et al., 1995, Downey and Fellows, 1995b]. Just as NP-hardness is used as evidence that a problem probably is not polynomial time solvable, there exists a hierarchy of complexity classes above FPT, and showing that a parameterized problem is hard for one of these classes is considered evidence that the problem is unlikely to be fixed-parameter tractable. Indeed, assuming the Exponential Time Hypothesis, a problem hard for  $W[1]$  does not belong to FPT [Downey and Fellows, 1999a]. The main classes in this hierarchy are:

$$\text{FPT} \subseteq W[1] \subseteq W[2] \subseteq \dots \subseteq W[P] \subseteq \text{XP},$$

where a parameterized problem belongs to the class  $XP$  if there exists an algorithm for it with running time bounded by  $n^{g(k)}$ . A parameterized problem is said to be para-NP-complete if it is NP-complete even for constant values of the parameter. A classic example of a para-NP-complete problem is graph coloring parameterized by the number of colors — recall that it is NP-complete to determine if a graph can be properly colored with three colors. Observe that a para-NP-complete problem does not belong to  $XP$  unless  $P = NP$ . We are now ready to describe our contributions in greater detail.

## 1.1 Our Contributions

We consider three standard approval scoring mechanisms, namely the *minisum*, *approval*, and *net approval scores*. The last two scores, as originally defined in the literature, are designed to simply give us the total amount of approval that a committee incurs from all the voters. The scores themselves are therefore non-decreasing functions of  $n$ , and as such, the question of outliers is not interesting if we use the scores directly (in particular, it is impossible to improve these scores by removing votes). Therefore, we consider the dual scoring system that complements the original — namely, we score a committee based on the amount *disapproval* that it incurs from all the votes, and seek to minimize the *total disapproval*. Typically, for any notion of approval, there are either one or two natural complementary notions of disapproval that present themselves (discussed in greater detail below). This formulation is consistent with the idea that we want our scores to capture “distance” rather than closeness. We note that in terms of scores, these rules are equivalent to the original, but choosing to ask the minimization question allows us to formulate the problem of finding the best committee in the presence of outliers.

**Remark 1.** *One might also consider the approval score per vote instead of the total score, as described earlier. However, we chose to use the notion of disapproval because of its consistency with the other distance-based rules (like minisum and minimax). All the variations are equivalent as scoring functions, and we note that our choice is only a matter of exposition.*

In Table 1, we summarize the notions of distances between a committee and a vote. Each of these notions naturally gives rise to a score-based voting rule. Formally, for any distance function  $s : 2^C \times 2^C \rightarrow \mathbb{N}$  between two subsets of candidates, we overload notation and define the corresponding score function  $s : 2^C \times \mathcal{V} \rightarrow \mathbb{N}$

Measures of approval		Measures of disapproval	
Minisum	$ \mathcal{X} \Delta \mathcal{S} $		Minisum
Approval	$ \mathcal{X} \cap \mathcal{S} $	$ \mathcal{X} \setminus \mathcal{S} $	Disapproval
		$ \mathcal{S} \setminus \mathcal{X} $	
Net Approval	$ \mathcal{X} \cap \mathcal{S}  -  \mathcal{X} \setminus \mathcal{S} $	$ \mathcal{X} \setminus \mathcal{S}  -  \mathcal{X} \cap \mathcal{S} $	Net Disapproval

Table 1: The approval-based rules that are considered in this paper. Here,  $\mathcal{X}$  is a candidate committee and  $\mathcal{S}$  is a vote, and the table illustrates the possible scores that  $\mathcal{X}$  can incur from  $\mathcal{S}$ . The total score of  $\mathcal{X}$  will be the sum of these scores taken over  $\mathcal{S}$ .

for a set of votes  $\mathcal{V} = \{\mathcal{S}_1, \dots, \mathcal{S}_n\}$  as follows:

$$s(\mathcal{X}, \mathcal{V}) := \sum_{i=1}^n s(\mathcal{X}, \mathcal{S}_i)$$

For the *winner determination* problem, the goal is to find a committee  $\mathcal{X}$  of size  $m^*$  that minimizes  $s(\mathcal{X}, \mathcal{V})$ . For all the scoring rules in Table 1, a winning committee of size  $m^*$  can be found in polynomial time for any  $m^* \leq m$ . We refer the reader to Table 2 for an overview of the notation we use in this paper. We are now ready to define the problem of winner determination for a scoring rule  $s$  in the presence of outliers:

**$s$ -OUTLIERS**

**Input:** A set of votes  $\mathcal{V} = \{\mathcal{S}_1, \dots, \mathcal{S}_n\}$  over a set of candidates  $\mathcal{C} = \{c_1, \dots, c_m\}$ , a committee size  $m^*$ , a target number of non-outliers  $n^*$ , and a target score  $t$ .

**Question:** Does there exist a committee  $\mathcal{C}^* \subset \mathcal{C}$  and a set of non-outliers  $\mathcal{V}^* \subset \mathcal{V}$  such that  $|\mathcal{V}^*| \geq n^*$ ,  $|\mathcal{C}^*| = m^*$ , and  $s(\mathcal{V}^*, \mathcal{C}^*) \leq t$ ?

**Remark 2.** We will focus on only one variant of disapproval for the approval scoring rule, namely the one given by the top row in Table 1. The other variation is symmetric and our results hold in the exact same fashion, as explained in Lemma 9 (in Section 6).

We first show that the  $s$ -OUTLIERS problem is NP-complete for all the scoring rules considered here, even in the special case when every vote approves exactly three candidates *and* every candidate is approved by exactly three votes. This also establishes the para-NP-hardness of the problem with respect to the (combined) parameters (*maximum*) *number of candidates approved by any vote* and (*maximum*) *number of votes that approve a candidate*.

**Theorem 1.** Let  $s \in \{\text{Minisum}, \text{Disapproval}, \text{Net-Disapproval}\}$ . The  $s$ -OUTLIERS problem is NP-complete even if every vote approves exactly 3 candidates and every candidate is approved by exactly 3 votes.

To initiate the parameterized study of the  $s$ -OUTLIERS problem, we propose the following parameters: the size of the committee ( $m^*$ ), the number of candidates *not* in the committee ( $\overline{m}$ ), the number of non-outliers ( $n^*$ ), the number of outliers ( $\overline{n}$ ), and the target score  $t$  (which one might think of as the “solution size” or the *standard parameter*).

Our main results are the following. For any subset of these parameters, we establish if the  $s$ -OUTLIERS problem is FPT or W[1]-hard when  $s \in \{\text{Minisum}, \text{Net Approval}\}$ , and for the Disapproval voting rule, we have a classification for all cases but one. Specifically, for the minisum voting rule, we establish the following dichotomy.

**Theorem 2.** Let  $\mathcal{P} = \{m^*, \overline{m}, n^*, \overline{n}, t\}$ , and  $\mathcal{Q} \subseteq \mathcal{P}$ . The MINISUM-OUTLIERS problem parameterized by  $\mathcal{Q}$  is FPT if  $\mathcal{Q}$  contains either  $\{m^*, \overline{m}\}$ ,  $\{n^*, \overline{n}\}$ , or  $\{\overline{n}, t\}$ , and is W[1]-hard otherwise.

Number of votes	$n$	Set of candidates	$\mathcal{C}$
Number of candidates	$m$	Committee chosen	$\mathcal{C}^*$
Number of outliers	$\bar{n}$	Non-committee	$\bar{\mathcal{C}} (= \mathcal{C} \setminus \mathcal{C}^*)$
Number of non-outliers	$n^* (= n - \bar{n})$	Set of votes	$\mathcal{V}$
Size of committee	$m^*$	Set of non-outliers	$\mathcal{V}^*$
Size of non-committee	$\bar{m} (= m - m^*)$	Set of outliers	$\bar{\mathcal{V}} (= \mathcal{V} \setminus \mathcal{V}^*)$
Score of the committee	$t$	Hamming distance	$h(\cdot)$

Table 2: Notation table.

For the net disapproval voting rule, we establish the following dichotomy.

**Theorem 3.** *Let  $\mathcal{P} = \{m^*, \bar{m}, n^*, \bar{n}, t\}$ , and  $\mathcal{Q} \subseteq \mathcal{P}$ . The NET-DISAPPROVAL-OUTLIERS problem parameterized by  $\mathcal{Q}$  is FPT if  $\mathcal{Q}$  contains either  $\{m^*, \bar{m}\}$  or  $\{n^*, \bar{n}\}$ , and is W[1]-hard otherwise.*

Further, for the disapproval voting rule, we have the following theorem that classifies all cases but one.

**Theorem 4.** *Let  $\mathcal{P} = \{m^*, \bar{m}, n^*, \bar{n}, t\}$ , and  $\mathcal{Q} \subseteq \mathcal{P}$ . The DISAPPROVAL-OUTLIERS problem parameterized by  $\mathcal{Q}$  is FPT if  $\mathcal{Q}$  contains either  $\{m^*, \bar{m}\}$  or  $\{n^*, \bar{n}\}$ , and is W[1]-hard for all other cases except for  $\mathcal{Q} = \{\bar{m}, \bar{n}, t\}$ .*

Apart from these classification results in the context of exact algorithms, we also pursue approximation algorithms and other special cases. We briefly summarize our contributions in these contexts below.

*Approximation Results.* We provide a polynomial time  $\varepsilon \bar{m}$ -approximation algorithm for optimizing the minisum score considering outliers, for every constant  $\varepsilon > 0$  [Theorem 5]. We also show a  $(1 + \varepsilon)$ -approximation algorithm for optimizing the minisum score considering outliers running in time  $n^{O(\log m / \varepsilon^2)}$ , for every constant  $\varepsilon > 0$  [Theorem 6]. On the hardness side, we show that in the presence of outliers, there does not exist an  $\alpha(m, n)$ -approximation algorithm for optimizing score of the selected committee for any computable function  $\alpha(\cdot, \cdot)$ , for the net disapproval [Corollary 2] and disapproval voting rules (unless  $P = NP$ ) [Lemma 8].

*Other Special Cases.* We show that when every voter approves at most one candidate, then the MINISUM-OUTLIER problem can be solved in polynomial amount of time. For some W[1]-hard cases, we show that the problem becomes FPT if the maximum number of candidates approved by a vote or the maximum number of votes that approve of a candidate is a constant. We refer the reader to Section 6 for a more detailed discussion.

## 1.2 Related Work

The notion of outliers is quite prominent in the literature pertaining to closest strings, where the usual setting is that we are given  $n$  strings over some alphabet  $\Sigma$ , and the task is to find a string  $x$  that minimizes either the maximum Hamming distance from any string, or the sum of

its Hamming distances from all strings. In the context of social choice theory, the notion of outliers are intimately related to manipulation and control of election by removing voters and candidates [Bartholdi et al., 1992, Bartholdi III et al., 1989b, Conitzer et al., 2007, Chevalerey et al., 2007, Hemaspaandra et al., 2007, Brandt et al., 2012, Faliszewski et al., 2009, Faliszewski et al., 2011, Meir et al., 2008, Procaccia et al., 2007, Elkind et al., 2009, Betzler and Uhlmann, 2008].

In the minimax notion of distance, the problem of finding a closest string in the presence of outliers was initiated in [Boucher and Ma, 2011], and was further refined in [Boucher et al., 2013]. The work in [Boucher et al., 2013] demonstrates hardness of approximation for variations involving outliers. With the minisum notion of distance, the study of finding the best string with respect to outliers was studied in [Lo et al., 2014]. In contrast with the minimax rule, this work shows a PTAS for *Consensus Sequence with Outliers*.

The fact that the closest string problem has similarities to the setup of approval ballots has been explored in the past, see, for instance [Byrka and Sornat, 2014]. More specifically, and again in the spirit of observing similarities with the closest string family of problems, the minimax approval rule has been studied from the perspective of outliers [Misra et al., 2015]. To the best of our knowledge, our work is the first comprehensive study of the complexity behavior of approval ballots in the presence of outliers.

## 2 Preliminaries

In this section we briefly recall some terminology and notation that we will use throughout.

**Approval Voting** Let  $\mathcal{V} = \{\mathcal{S}_1, \dots, \mathcal{S}_n\}$  be the set of all *votes* and  $\mathcal{C} = \{c_1, \dots, c_m\}$  the set of all *candidates*. Each vote  $\mathcal{S}_i$  is a subset of  $\mathcal{C}$ . We say voter  $i$  approves a candidate  $x \in \mathcal{C}$  if  $x \in \mathcal{S}_i$ ; otherwise we say, voter  $i$  does not approve the candidate  $x$ . A voting rule  $r$  for choosing a committee of size  $m^*$  is a mapping  $r : \cup_{n \geq 1} (2^{\mathcal{C}})^n \rightarrow \mathbb{P}_{m^*}(\mathcal{C})$ , where  $\mathbb{P}_{m^*}(\mathcal{C})$  is the set of all subsets of  $\mathcal{C}$  of size  $m^*$ . We call a voting rule  $r$  a *scoring rule* if there exists a scoring function  $s : \cup_{n \geq 1} (2^{\mathcal{C}})^n \times 2^{\mathcal{C}} \rightarrow \mathbb{N}$  such that  $r(\mathcal{V}) = \operatorname{argmin}_{X \in \mathbb{P}_{m^*}(\mathcal{C})} s(\mathcal{V}, X)$ , for every  $\mathcal{V} \in (2^{\mathcal{C}})^n$ . We refer to Table 1 for some common scoring rules. Unless mentioned otherwise, we use the notations listed in Table 2.

**Algorithmic Terminology** Given a set  $A$ , we denote the complement of  $A$  by  $A^c$ . For a positive integer  $\ell$ , we denote the set  $\{1, 2, \dots, \ell\}$  by  $[\ell]$ . We use the notation  $O^*(f(m, n))$  to denote  $O(f(m, n) \operatorname{poly}(m, n))$ .

For a minimization problem  $\mathcal{P}$ , we say an algorithm  $\mathcal{A}$  archives an approximation factor of  $\alpha$  if  $\mathcal{A}(\mathcal{I}) \leq \alpha \operatorname{OPT}(\mathcal{I})$  for every problem instance  $\mathcal{I}$  of  $\mathcal{P}$ . In the above,  $\operatorname{OPT}(\mathcal{I})$  denotes the value of the optimal solution of the problem instance  $\mathcal{I}$ .

**Definition 1. (Parameterized Reduction [Downey and Fellows, 1999b])**

Let  $\Gamma_1$  and  $\Gamma_2$  be parameterized problems. A parameterized reduction from  $\Gamma_1$  to  $\Gamma_2$  is an algorithm that, given an instance  $(x, k)$  of  $\Gamma_1$ , outputs an instance  $(x', k')$  of  $\Gamma_2$  such that:

1.  $(x, k)$  is a yes-instance of  $\Gamma_1$  if and only if  $(x', k')$  is a yes-instance of  $\Gamma_2$ ,
2.  $k' \leq g(k)$  for some computable function  $g$ , and
3. the running time of the algorithm is  $f(k)|x|^{O(1)}$  for some computable function  $f$ .

## 3 Classical Complexity: NP-hardness Results

We begin by showing that even for rules where winner determination is polynomial time solvable, the possibility of choosing outliers makes the winner determination problem significantly harder. In particular,

we show that the  $s$ -OUTLIERS problem is hard for minisum, disapproval, and net disapproval voting rules. We reduce from the well known VERTEX COVER problem, which is known to be NP-complete even on 3-regular graphs [Garey and Johnson, 1979].

**3-VERTEX-COVER**

**Input:** A 3-regular graph  $G$  and a positive integer  $k$ .

**Question:** Is there a subset  $S$  of at most  $k$  vertices such that  $G[V \setminus S]$  is an independent set?

**Theorem 1.** *Let  $s \in \{\text{Minisum}, \text{Disapproval}, \text{Net-Disapproval}\}$ . The  $s$ -OUTLIERS problem is NP-complete even if every vote approves exactly 3 candidates and every candidate is approved by exactly 3 votes.*

*Proof.* First let us prove the result for the minisum voting rule. The problem is clearly in NP. To prove NP-hardness, we reduce the 3-VERTEX-COVER problem to the MINISUM-OUTLIERS. Let  $(\mathcal{G} = (U, E), k)$  be an arbitrary instance of 3-VERTEX-COVER problem. Let  $U = \{u_1, \dots, u_n\}$ . We will introduce a candidate and a vote corresponding to every vertex, and have candidate  $i$  approved by a vote  $j$  if and only if the vertices corresponding to  $i$  and  $j$  are adjacent in  $G$ . We define the corresponding instance  $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$  of the MINISUM-OUTLIERS problem as follows:

$$\mathcal{V} = \{\mathcal{S}_1, \dots, \mathcal{S}_n\}, \mathcal{C} = \{c_1, \dots, c_n\},$$

$$\mathcal{S}_i = \{c_j : (u_i, u_j) \in E\} \quad \forall i \in [n],$$

$$n^* = n - k, m^* = k, t = (n - k)(k - 3)$$

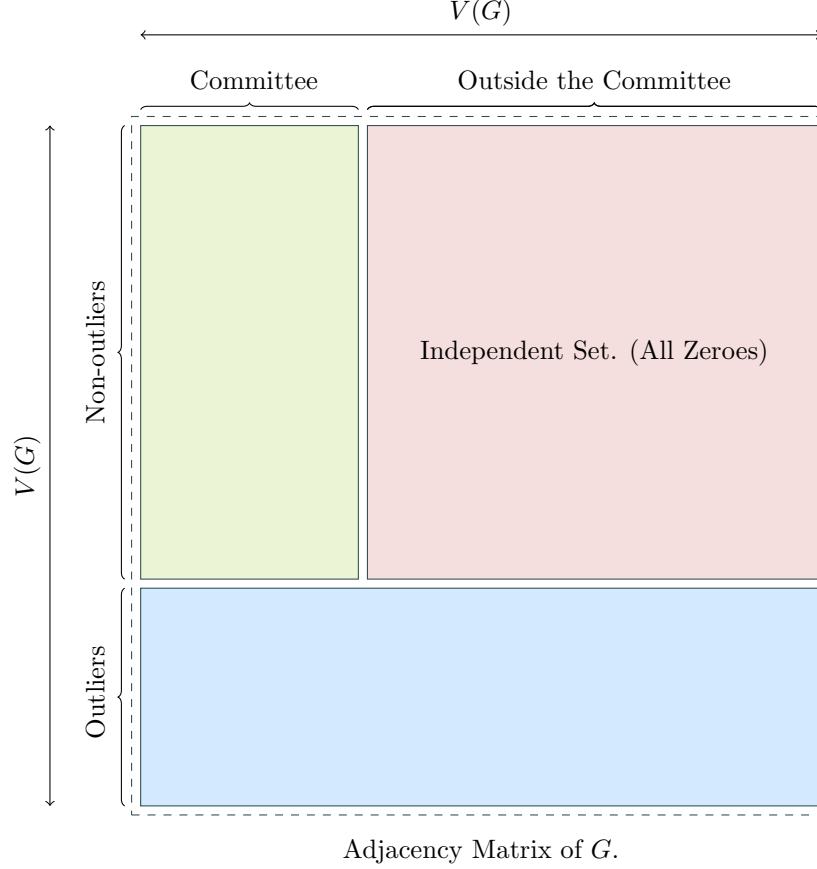
Notice that,  $|\mathcal{S}_i| = 3$  and each candidate is approved in exactly 3 votes since  $\mathcal{G}$  is 3-regular. We claim that the two instances equivalent. In the forward direction, suppose  $W \subset U$  forms a vertex cover with  $|W| = k$ . Consider the committee  $\mathcal{C}^* = \{c_i : u_i \in W\}$  and the set of nonoutliers  $\mathcal{V}^* = \{\mathcal{S}_i : u_i \notin W\}$ . We claim that  $h(\mathcal{S}_i, \mathcal{C}^*) = (k - 3)$  for every  $\mathcal{S}_i \in \mathcal{V}^*$  thereby proving  $h(\mathcal{V}^*, \mathcal{C}^*) = (n - k)(k - 3)$ . Notice that,  $h(\mathcal{S}_i, \mathcal{C}^*) \geq (k - 3)$  since  $|\mathcal{C}^*| = k$  and  $|\mathcal{S}_i| = 3$  for every  $\mathcal{S}_i \in \mathcal{V}^*$ . Suppose, there exist an  $\mathcal{S}_i \in \mathcal{V}^*$  such that  $h(\mathcal{S}_i, \mathcal{C}^*) < k - 3$ . Then there exist a candidate  $c_j \in \mathcal{S}_i \setminus \mathcal{C}^*$ . However, this implies that the edge  $(u_i, u_j) \in E$  is not covered by  $W$  which contradicts that fact that  $W$  is a vertex cover.

For the reverse direction, suppose there exists a set of outliers  $\overline{\mathcal{V}} \subset \mathcal{V}$  and a committee  $\mathcal{C}^* \subset \mathcal{C}$  such that  $|\overline{\mathcal{V}}| \leq \bar{n}$ ,  $|\mathcal{C}^*| = m^*$ , and  $h(\mathcal{V}^*, \mathcal{C}^*) \leq t$ . Since, adding votes in the set of outliers  $\overline{\mathcal{V}}$  can only reduce  $h(\mathcal{V}^*, \mathcal{C}^*)$ , we may assume without loss of generality that  $|\overline{\mathcal{V}}| = \bar{n}$ . Now, since  $h(\mathcal{S}_i, \mathcal{C}^*) \geq (k - 3)$  for every  $\mathcal{S}_i \in \mathcal{V}^*$ , we have  $h(\mathcal{V}^*, \mathcal{C}^*) = (n - k)(k - 3)$  and thus  $h(\mathcal{S}_i, \mathcal{C}^*) = (k - 3)$  for every  $\mathcal{S}_i \in \mathcal{V} \setminus \overline{\mathcal{V}}$ . Let  $\mathcal{X} = \{u_i \in U : \mathcal{S}_i \in \mathcal{V}^*\}$  and  $\mathcal{Y} = \{u_i \in U : c_i \in \mathcal{C}^*\}$ . We claim that  $\mathcal{X}$  covers all the edges incident on the vertices in  $\mathcal{Y}$ . Indeed, otherwise there exist a  $u_i \in \mathcal{X}$  such that all the edges incident on  $u_i$  are not covered by  $\mathcal{Y}$  and thus we have  $h(\mathcal{S}_i, \mathcal{C}^*) > k - 3$ . We now construct a vertex cover  $\mathcal{W}$  of  $\mathcal{G}$  of size at most  $k$  as follows. If  $\mathcal{X}$  and  $\mathcal{Y}$  are disjoint then,  $\mathcal{Y}$  forms a vertex cover of  $\mathcal{G}$ . Otherwise we let  $\mathcal{Y}_0 = \mathcal{Y}, u \in \mathcal{X} \cap \mathcal{Y}_0, v \in (\mathcal{U} \setminus \mathcal{X}) \setminus \mathcal{Y}_0$  and notice that  $\mathcal{Y}_0 \setminus \{u\}$  still covers all the edges incident on  $\mathcal{X}$ . We define  $\mathcal{Y}_1 = (\mathcal{Y}_0 \setminus \{u\}) \cup \{v\}$ . We note that  $|\mathcal{X} \cap \mathcal{Y}_1| = |\mathcal{X} \cap \mathcal{Y}_0| - 1$  and  $|\mathcal{Y}_1| = |\mathcal{Y}_0|$ . We iterate this process for  $|\mathcal{X} \cap \mathcal{Y}|$  many times. Let us call  $\mathcal{Y}_{|\mathcal{X} \cap \mathcal{Y}|}$   $\mathcal{W}$ . By the argument above,  $\mathcal{W}$  covers all the edges incident on the vertices in  $\mathcal{X}$ ,  $\mathcal{W} \cap \mathcal{X} = \emptyset$ , and thus we have  $\mathcal{W} = \mathcal{U} \setminus \mathcal{X}$ . Hence,  $\mathcal{W}$  forms a set cover of  $\mathcal{G}$ .

The proofs for the other rules are identical, except for the values of the target score. We define  $t = (n - k)(k - 3)$  for the disapproval, and  $t = (n - k)(k - 6)$  for the net disapproval voting rule. It is easily checked that the details are analogous.  $\square$

Theorem 1 immediately yields the following corollary.

**Corollary 1.** *Let  $s \in \{\text{Minisum}, \text{Disapproval}, \text{Net-Disapproval}\}$ . Then the  $s$ -OUTLIERS problem is para-NP-hard when parameterized by the sum of the maximum number of approvals that a candidate obtains and the maximum number of candidates that a vote approves.*



## 4 Parameterized Complexity Results

In this section, we present the results on parameterized complexity of the  $s$ -OUTLIERS problem.

### 4.1 FPT algorithms

The following result follows from the fact that the  $s$ -OUTLIERS problem is polynomial time solvable for all the voting rules considered here if we know either the committee or the non-outliers of the solution.

**Proposition 1.** *Let  $s \in \{\text{Minisum}, \text{Disapproval}, \text{Net-Disapproval}\}$ . Then, there is a  $O^*(2^m)$  time algorithm and a  $O^*(2^n)$  time algorithm for the  $s$ -OUTLIERS problem.*

Now, we show that the MINISUM-OUTLIERS problem, parameterized by  $(t + \bar{n})$ , is FPT.

**Lemma 1.** *There is a  $O^*(2^{t+\bar{n}})$  time algorithm for the MINISUM-OUTLIERS problem.*

*Proof.* We consider the following two cases.

**Case  $t < n - \bar{n}$ :** In this case there exist a vote say  $\mathcal{S}_i$  such that  $h(\mathcal{X}, \mathcal{S}_i) = 0$ , where  $\mathcal{X}$  is the committee selected. Hence, we can iterate over all possible such  $\mathcal{S}_i$  (if  $|\mathcal{S}_i| \neq m^*$  then, do not consider that  $\mathcal{S}_i$ ) and fix  $\mathcal{X}$  to be  $\mathcal{S}_i$ . We now identify the non-outliers  $\mathcal{V}^*$  with respect to the committee  $\mathcal{X}$ .

**Case  $t \geq n - \bar{n}$ :** Use the algorithm in Proposition 1 which runs in  $O^*(2^n) = O^*(2^{t+\bar{n}})$  time. □



## 4.2 W[1]-hardness Results

In this section, we establish our W[1]-hardness results. To begin with, we focus on parameters combined with the target score  $t$ . Notice that we have tractability when we combine either  $(\bar{n} + n^*)$  or  $(\bar{m} + m^*)$  along with  $t$  (follows from Proposition 1). For minisum, we even have tractability for  $t$  and  $\bar{n}$ , from Lemma 1. Therefore, the interesting combinations for the minisum rule are  $(t + n^* + \bar{m})$  and  $(t + n^* + m^*)$ . We first consider the combination  $(t + n^* + \bar{m})$ , and show W[1]-hardness for all the voting rules considered here by exhibiting a parameterized reduction from the  $d$ -CLIQUE problem which is known to be W[1]-hard parameterized by clique size [Downey and Fellows, 2013].

<i>d</i> -CLIQUE	<b>Parameter:</b> $k$
<b>Input:</b> A $d$ -regular graph $G$ and a positive integer $k$ .	
<b>Question:</b> Is there a clique of size $k$ ?	

**Lemma 2.** *The MINISUM-OUTLIERS problem is W[1]-hard, when parameterized by  $(t + n^* + \bar{m})$ . Also, for  $s \in \{\text{disapproval}, \text{net disapproval}\}$ , the  $s$ -OUTLIERS problem is W[1]-hard when parameterized by  $(n^* + \bar{m})$ , even when  $t = 0$ .*

*Proof.* First let us prove the result for the MINISUM-OUTLIERS problem. We exhibit a parameterized reduction from the  $d$ -CLIQUE problem to the MINISUM-OUTLIERS problem thereby proving the result. Let  $(\mathcal{G} = (U, E), k)$  be an arbitrary instance of the  $d$ -CLIQUE problem. Let  $U = \{u_1, \dots, u_n\}$  and  $E = \{e_1, \dots, e_m\}$ . We define the corresponding instance  $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$  of the MINISUM-OUTLIERS problem as follows:

$$\mathcal{V} = \{\mathcal{S}_1, \dots, \mathcal{S}_m\}, \mathcal{C} = \{c_1, \dots, c_n\}, \mathcal{S}_i = \{c_j : u_j \notin e_i\} \ \forall i \in [n],$$

$$n^* = \binom{k}{2}, \bar{m} = k, t = (k-2) \binom{k}{2}$$

We claim that the two instances are equivalent. In the forward direction, suppose  $W \subset U$  forms a clique with  $|W| = k$ , and let  $Q$  denote the set of edges that have both endpoints in  $W$ . Then the committee  $\mathcal{C}^* = \{c_i : u_i \notin W\}$  along with the set of non-outliers  $\mathcal{V}^* = \{\mathcal{S}_i : u_i \in Q\}$  achieves minisum score of  $(k-2) \binom{k}{2}$ .

In the reverse direction, suppose there exist a set of non-outliers  $\mathcal{V}^* \subset \mathcal{V}$  and a committee  $\mathcal{C}^* \subset \mathcal{C}$  such that  $h(\mathcal{V}^*, \mathcal{C}^*) \leq (k-2) \binom{k}{2}$ . We claim that the vertices in  $W = \{u_i : c_i \in \mathcal{C}^*\}$  form a clique with edge set  $Q = \{e_i : \mathcal{S}_i \in \mathcal{V}^*\}$ . If not then, there exist a vote  $\mathcal{S}_j \in \mathcal{V}^*$  such that  $h(\mathcal{S}_j, \mathcal{C}) \geq k$ . On the other hand we have  $h(\mathcal{S}_i, \mathcal{C}) \geq k-2$  for every  $\mathcal{S}_i \in \mathcal{V}^*$ . This implies that  $h(\mathcal{V}^*, \mathcal{C}^*) > (k-2) \binom{k}{2}$ , which is a contradiction.

We define  $t = 0$  for the disapproval and net disapproval voting rules. It is easily checked that the details are analogous.  $\square$

We now turn to the combination  $(m^* + n^* + t)$ . Again, we have W[1]-hardness for all voting rules considered.

**Lemma 3.** *Let  $s \in \{\text{Minisum}, \text{Net-Disapproval}, \text{Disapproval}\}$ . Then the  $s$ -OUTLIERS problem is W[1]-hard, when parameterized by  $(t + m^* + n^*)$ , even when every vote approves exactly two candidates.*

We now consider the other combinations involving  $t$  that are non-trivial for voting rules different from minisum, namely  $(t + \bar{n} + m^*)$  and  $(t + \bar{n} + \bar{m})$ . Lemma 4 below shows W[1]-hardness for the first combination, for the Disapproval voting rule. This also shows W[1]-hardness with respect to the parameter  $(\bar{n} + m^*)$  alone, for the minisum voting rule.

**Lemma 4.** *The MINISUM-OUTLIERS problem is W[1]-hard when parameterized by  $(\bar{n} + m^*)$ . Also, the DISAPPROVAL-OUTLIERS problem is W[1]-hard parameterized by  $(m^* + \bar{n})$ , even when the target score  $t = 0$ . In particular,  $s$ -OUTLIERS is W[1]-hard parameterized by  $(m^* + \bar{n} + t)$ .*

*Proof.* To begin with, let us prove the result for the MINISUM-OUTLIERS problem. We exhibit a parameterized reduction from the  $d$ -CLIQUE problem to the MINISUM-OUTLIERS problem thereby proving the result. Let  $(\mathcal{G} = (U, E), k)$  be an arbitrary instance of the  $d$ -CLIQUE problem. Let  $U = \{u_1, \dots, u_n\}$  and  $E = \{e_1, \dots, e_m\}$ . We define the corresponding instance  $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$  of the MINISUM-OUTLIERS problem as follows:

$$\mathcal{V} = \{\mathcal{S}_1, \dots, \mathcal{S}_n\}, \mathcal{C} = \{c_1, \dots, c_m\}, \mathcal{S}_i = \{c_j : u_i \notin e_j\} \quad \forall i \in [n],$$

$$\bar{n} = k, m^* = \binom{k}{2}, t = (m - \binom{k}{2})(n - k)$$

We claim that the two instances are equivalent. In the forward direction, suppose  $W \subset U$  forms a clique with  $|W| = k$ , and let  $Q$  denote the set of edges that have both endpoints in  $W$ . Then the committee  $\mathcal{C}^* = \{c_i : e_i \in Q\}$  and the set of non-outliers  $\mathcal{V}^* = \{\mathcal{S}_i : u_i \notin W\}$  achieves minisum score of  $(m - \binom{k}{2})(n - k)$ .

In the reverse direction, suppose there exist a set of outliers  $\bar{\mathcal{V}} \subset \mathcal{V}$  and a committee  $\mathcal{C}^* \subset \mathcal{C}$  such that  $h(\mathcal{V}^*, \mathcal{C}^*) \leq (m - \binom{k}{2})(n - k)$ . We claim that the vertices in  $W = \{u_i : \mathcal{S}_i \in \bar{\mathcal{V}}\}$  form a clique. If not then, there exist a candidate  $x \in \mathcal{C}^*$  that is not approved by at least one vote in  $\mathcal{V}^*$ . However, this implies that  $h(\mathcal{V}^*, \mathcal{C}^*) > (m - \binom{k}{2})(n - k)$ , which is a contradiction.

We define  $t = 0$  for the disapproval voting rule. It is easily checked that the details are analogous.  $\square$

Next, we show that NET-DISAPPROVAL-OUTLIERS is  $W[1]$ -hard with respect to the combined parameter  $(\bar{n} + t + m^*)$ .

**Lemma 5.** *The NET-DISAPPROVAL-OUTLIERS problem is  $W[1]$ -hard, when parameterized by  $(\bar{n} + m^*)$ , even when the target score is 0. In particular, NET-DISAPPROVAL-OUTLIERS problem is  $W[1]$ -hard, when parameterized by  $(t + \bar{n} + m^*)$ .*

*Proof.* We reduce the CLIQUE problem to the NET-APPROVAL-OUTLIER problem. Let  $(\mathcal{G} = (U, E), k)$  be an arbitrary instance of the CLIQUE problem. Let  $U = \{u_1, \dots, u_n\}$ . We define the corresponding instance  $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$  of the NET-APPROVAL-OUTLIER problem as follows:

$$\mathcal{C} = \{c_e : e \in E\} \cup D, \text{ where } |D| = \binom{k}{2},$$

$$\mathcal{V} = \{\mathcal{S}_1, \dots, \mathcal{S}_n\} \cup \{\mathcal{T}_1, \dots, \mathcal{T}_{n+2k}\},$$

$$\mathcal{S}_i = \{c_e : u_i \notin e\}, \forall i \in [n], \quad \mathcal{T}_j = D, \forall j \in \left[\binom{k}{2}\right],$$

$$\bar{n} = k, m^* = 2\binom{k}{2}, t = 0$$

We claim that the two instances are equivalent. In the forward direction, suppose  $\mathcal{G}$  has a clique on a subset of vertices  $W \subset U$  of size  $k$ ; let  $Q$  be the clique edges. Then we define the set of outlier votes to be  $\bar{\mathcal{V}} = \{\mathcal{S}_i : u_i \in W\}$  and the committee to be the set of candidates  $\mathcal{C}^* = D \cup \{c_e : e \in Q\}$ . Now we have  $\sum_{\mathcal{S} \in \mathcal{V}^*} |\mathcal{S} \cap \mathcal{C}^*| = (n - k)\binom{k}{2} + (n + 2k)|D| = (n + k)\binom{k}{2}$  and  $\sum_{\mathcal{S} \in \mathcal{V}^*} |\mathcal{S}^c \cap \mathcal{C}^*| = (n - k)|D| + (n + 2k)\binom{k}{2} = (n + k)\binom{k}{2}$  thereby achieving the net approval score  $t$  of 0.

In the reverse direction, suppose we have a set of outlier votes  $\bar{\mathcal{V}}$  and a committee  $\mathcal{C}^*$  which achieves a net approval score  $t$  of 0. First observe that we can assume without loss of generality that  $D$  is a subset of  $\mathcal{C}^*$  since irrespective of the outliers chosen, every candidate in  $D$  receives at least  $(n + k)$  approvals and every candidate not in  $D$  receives at most  $n$  approvals. Now since the committee  $\mathcal{C}^*$  contains  $D$ , for every  $j \in [n + 2k]$ , the vote  $\mathcal{T}_j$  contributes at most  $|D| - \binom{k}{2} = 0$  to the net approval score, whereas, for every  $i \in [n]$ , the vote  $\mathcal{S}_i$  contributes at least  $\binom{k}{2} - |D| = 0$  to the net approval score. Hence, we may assume without loss of generality that  $\mathcal{T}_j$  belongs to the set of non-outliers  $\mathcal{V}^*$  for every  $j \in [n + 2k]$ . Now we claim

that the set of edges  $Q = \{e : c_e \in \mathcal{C}^*\}$  must form a clique on the set of vertices  $W = \{u : \mathcal{S}_u \in \overline{\mathcal{V}}\}$ . If not then,  $\sum_{\mathcal{S} \in \mathcal{V}^*} |\mathcal{S} \cap \mathcal{C}^*| < (n+k) \binom{k}{2}$  and  $\sum_{\mathcal{S} \in \mathcal{V}^*} |\mathcal{S}^c \cap \mathcal{C}^*| > (n+k) \binom{k}{2}$  thereby making the net approval score  $t$  strictly more than 0.  $\square$

Finally, we show that for all rules considered, the  $s$ -OUTLIERS problem is  $W[1]$ -hard when parameterized by  $(\overline{n} + \overline{m})$ . In particular, for the NET-DISAPPROVAL-OUTLIERS problem, we have  $W[1]$ -hardness even when  $t = 0$  and every candidate is approved in exactly two votes.

**Lemma 6.** *Let  $s \in \{\text{Minisum}, \text{Disapproval}, \text{Net-Disapproval}\}$ . Then  $s$ -OUTLIERS is  $W[1]$ -hard, when parameterized by  $(\overline{n} + \overline{m})$ , even when every candidate is approved in exactly two votes. Further, the NET-DISAPPROVAL-OUTLIERS problem, parameterized by  $(\overline{n} + \overline{m})$ , is  $W[1]$ -hard even when  $t = 0$  and every candidate is approved in exactly two votes. In particular, the NET-DISAPPROVAL-OUTLIERS problem is  $W[1]$ -hard parameterized by  $(\overline{n} + \overline{m} + t)$ .*

*Proof.* First let us prove the result for the MINISUM-OUTLIERS problem. We exhibit a parameterized reduction from the  $d$ -CLIQUE problem to the MINISUM-OUTLIERS problem thereby proving the result. Let  $(\mathcal{G} = (U, E), k)$  be an arbitrary instance of the  $d$ -CLIQUE problem. Let  $U = \{u_1, \dots, u_n\}$  and  $E = \{e_1, \dots, e_m\}$ . We define the corresponding instance  $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$  of the MINISUM-OUTLIERS problem as follows:

$$\mathcal{V} = \{\mathcal{S}_1, \dots, \mathcal{S}_n\}, \mathcal{C} = \{c_1, \dots, c_m\}, \mathcal{S}_i = \{c_j : u_i \in e_j\}, \forall i \in [n],$$

$$\overline{n} = k, \overline{m} = \binom{k}{2}, t = (n - k) \left( m - \binom{k}{2} - d \right)$$

We claim that the two instances are equivalent. In the forward direction, suppose  $W \subset U$  forms a clique with  $|W| = k$ , and let  $Q$  denotes the set of edges that have both endpoints in  $W$ . Consider the committee  $\mathcal{C}^* = \{c_i : e_i \notin Q\}$  and the set of outliers  $\overline{\mathcal{V}} = \{\mathcal{S}_i : u_i \in W\}$ . Consider now a vote  $\mathcal{S}_j \in \mathcal{V}^*$ . Note that  $|\mathcal{C}^* \cap \mathcal{S}_j| = d$ , since every edge incident on  $u_j$  is an edge that does not belong to  $Q$ , by the definition of  $\mathcal{C}^*$ . Further, this also implies that  $|\mathcal{C}^* \setminus \mathcal{S}_j| = \left( m - \binom{k}{2} - d \right)$ . Therefore,  $h(\mathcal{S}_j, \mathcal{C}^*) = \left( m - \binom{k}{2} - d \right)$ . Hence we have  $h(\mathcal{V}^*, \mathcal{C}^*) = (n - k) \left( m - \binom{k}{2} - d \right)$ .

For the reverse direction, suppose there exist a set of non-outliers  $\mathcal{V}^* \subset \mathcal{V}$  and a committee  $\mathcal{C}^* \subset \mathcal{C}$  such that  $h(\mathcal{V}^*, \mathcal{C}^*) \leq (n - k) \left( m - \binom{k}{2} - d \right)$ . We claim that the vertices in  $W = \{u_i : \mathcal{S}_i \in \overline{\mathcal{V}}\}$  form a clique. If not then there exists a vote  $\mathcal{S} \in \mathcal{V}^*$ , such that  $h(\mathcal{S}, \mathcal{C}^*) > \left( m - \binom{k}{2} - d \right)$ . However, for every vote  $\mathcal{S}' \in \mathcal{V}^*$ , we have  $h(\mathcal{S}', \mathcal{C}^*) \geq \left( m - \binom{k}{2} - d \right)$ . This makes  $h(\mathcal{V}^*, \mathcal{C}^*) > (n - k) \left( m - \binom{k}{2} - d \right)$  which is a contradiction.

The proofs for the other rules are identical, except for the values of the target score. We define  $t = (n - k) \left( m - \binom{k}{2} - d \right)$  for the disapproval voting rule. For the net disapproval voting rule, we add  $(m - 2d)$  many dummy candidates who are approved by every vote. We keep  $\overline{m} = \binom{k}{2}$  and make  $t = 0$ . It is easily checked that the remaining details are analogous.  $\square$

### 4.3 Proofs of the Main Theorems

We are now ready to present the proofs of Theorems 2 to 4. First, we recall the dichotomy for the minisum voting rule.

**Theorem 2.** *Let  $\mathcal{P} = \{m^*, \overline{m}, n^*, \overline{n}, t\}$ , and  $\mathcal{Q} \subseteq \mathcal{P}$ . The MINISUM-OUTLIERS problem parameterized by  $\mathcal{Q}$  is FPT if  $\mathcal{Q}$  contains either  $\{m^*, \overline{m}\}$ ,  $\{n^*, \overline{n}\}$ , or  $\{\overline{n}, t\}$ , and is  $W[1]$ -hard otherwise.*

*Proof.* Since  $m^* + \overline{m} = m$  and  $m^* + \overline{n} = n$ , the tractability results follow from Proposition 1 and Lemma 1. Now, we only have to consider subsets of parameters  $\mathcal{Q}$  such that:

1.  $\mathcal{Q}$  does not contain both  $m^*$  and  $\overline{m}$
2.  $\mathcal{Q}$  does not contain both  $n^*$  and  $\overline{n}$
3.  $\mathcal{Q}$  does not contain both  $t$  and  $\overline{n}$

Among such choices of  $\mathcal{Q}$ , we have the following cases.

1. Suppose  $t \in \mathcal{Q}$ . Then  $\mathcal{Q}$  is either a subset of  $\mathcal{Q}_1 = \{t, n^*, \overline{m}\}$  or a subset of  $\mathcal{Q}_2 = \{t, n^*, m^*\}$ . The hardness for all of these cases follow from Lemma 2 and Lemma 3, respectively.
2. Suppose  $\overline{n} \in \mathcal{Q}$ . Then  $\mathcal{Q}$  is either a subset of  $\mathcal{Q}_1 = \{\overline{n}, \overline{m}\}$  or a subset of  $\mathcal{Q}_2 = \{\overline{n}, m^*\}$ . The hardness for all of these cases follow from Lemma 4 and Lemma 6, respectively.
3. If neither  $t$  nor  $\overline{n}$  belongs to  $\mathcal{Q}$ , then  $\mathcal{Q}$  is either a subset of  $\mathcal{Q}_1 = \{n^*, m^*\}$ , or  $\mathcal{Q}_2 = \{n^*, \overline{m}\}$ . Note that these cases are already subsumed by Case (1) above.

This completes the proof of the theorem.  $\square$

Now, we turn to the case of the net disapproval voting rule.

**Theorem 3.** *Let  $\mathcal{P} = \{m^*, \overline{m}, n^*, \overline{n}, t\}$ , and  $\mathcal{Q} \subseteq \mathcal{P}$ . The NET-DISAPPROVAL-OUTLIERS problem parameterized by  $\mathcal{Q}$  is FPT if  $\mathcal{Q}$  contains either  $\{m^*, \overline{m}\}$  or  $\{n^*, \overline{n}\}$ , and is  $W[1]$ -hard otherwise.*

*Proof.* Since  $m^* + \overline{m} = m$  and  $m^* + \overline{n} = n$ , the tractability results follow from Proposition 1. Now, we only have to consider subsets of parameters  $\mathcal{Q}$  such that  $\mathcal{Q}$  does not contain both  $m^*$  and  $\overline{m}$ , and  $\mathcal{Q}$  does not contain both  $n^*$  and  $\overline{n}$ . Among such choices of  $\mathcal{Q}$ , we have the following cases.

1. Suppose  $n^* \in \mathcal{Q}$ . Then  $\mathcal{Q}$  is either a subset of  $\mathcal{Q}_1 = \{n^*, \overline{m}, t\}$  or a subset of  $\mathcal{Q}_2 = \{n^*, m^*, t\}$ . The hardness for all of these cases follow from Lemma 2 and Lemma 3, respectively.
2. Suppose  $\overline{n} \in \mathcal{Q}$ . Then  $\mathcal{Q}$  is either a subset of  $\mathcal{Q}_1 = \{\overline{n}, \overline{m}, t\}$  or a subset of  $\mathcal{Q}_2 = \{\overline{n}, m^*, t\}$ . The hardness for all these cases follow from Lemma 6 and Lemma 5, respectively.
3. If neither  $n^*$  nor  $\overline{n}$  belongs to  $\mathcal{Q}$ , then  $\mathcal{Q}$  is either a subset of  $\mathcal{Q}_1 = \{t, m^*\}$ , or  $\mathcal{Q}_2 = \{t, \overline{m}\}$ . Note that these cases are already subsumed by the cases above.

This completes the proof of the theorem.  $\square$

Finally, we turn to the case of the disapproval voting rules.

**Theorem 4.** *Let  $\mathcal{P} = \{m^*, \overline{m}, n^*, \overline{n}, t\}$ , and  $\mathcal{Q} \subseteq \mathcal{P}$ . The DISAPPROVAL-OUTLIERS problem parameterized by  $\mathcal{Q}$  is FPT if  $\mathcal{Q}$  contains either  $\{m^*, \overline{m}\}$  or  $\{n^*, \overline{n}\}$ , and is  $W[1]$ -hard for all other cases except for  $\mathcal{Q} = \{\overline{m}, \overline{n}, t\}$ .*

*Proof.* Since  $m^* + \overline{m} = m$  and  $m^* + \overline{n} = n$ , the tractability results follow from Proposition 1. Now, we only have to consider subsets of parameters  $\mathcal{Q}$  such that  $\mathcal{Q}$  does not contain both  $m^*$  and  $\overline{m}$ , and  $\mathcal{Q}$  does not contain both  $n^*$  and  $\overline{n}$ . Among such choices of  $\mathcal{Q}$ , we have the following cases.

1. Suppose  $n^* \in \mathcal{Q}$ . Then  $\mathcal{Q}$  is either a subset of  $\mathcal{Q}_1 = \{n^*, \overline{m}, t\}$  or a subset of  $\mathcal{Q}_2 = \{n^*, m^*, t\}$ . The hardness for all of these cases follows from Lemma 2 and Lemma 3, respectively.
2. Suppose  $\overline{n} \in \mathcal{Q}$ . Then  $\mathcal{Q}$  is either a subset of  $\mathcal{Q}_1 = \{\overline{n}, m^*, t\}$  or  $\mathcal{Q}_2 = \{\overline{n}, \overline{m}, t\}$ . The hardness for all subsets of  $\mathcal{Q}_1$  follows from Lemma 4. The status for  $\mathcal{Q}_2$ , as stated in the theorem, is open. The hardness for all strict subsets of  $\mathcal{Q}_2$  follows from Lemma 6 and the cases that are already resolved.
3. If neither  $n^*$  nor  $\overline{n}$  belongs to  $\mathcal{Q}$ , then  $\mathcal{Q}$  is either a subset of  $\mathcal{Q}_1 = \{t, m^*\}$ , or  $\mathcal{Q}_2 = \{t, \overline{m}\}$ . Note that these cases are already subsumed by the cases above.

This completes the proof of the theorem.  $\square$

## 5 Approximation Results

In this section, we describe our results in the context of approximation algorithms, where our goal is to minimize the target score, given a committee size  $m^*$  and a budget  $\bar{n}$  for the number of outliers as a part of input. In the first part, we focus on the minisum voting rule, and show an  $\varepsilon\bar{n}$  approximation algorithm, for every constant  $\varepsilon > 0$ . We also have a  $(1 + \varepsilon)$ -approximation algorithm whose running time is  $n^{O(\log m/\varepsilon^2)}$ , for every constant  $\varepsilon > 0$ . Subsequently, we show that for all the voting rules, the target score is **NP**-hard to approximate within any factor, as it is **NP**-hard already to determine if  $t = 0$ .

### 5.1 Approximation Algorithms

We now turn to some approximation algorithms for the MINISUM-OUTLIERS problem. Our first result is the following.

**Theorem 5.** *There is a  $\varepsilon\bar{n}$ -approximation algorithm for the MINISUM-OUTLIERS problem, for every constant  $\varepsilon > 0$ .*

Our next result is an approximation scheme with a subexponential running time, loosely based on the framework introduced in [Lo et al., 2014]. To this end, we will need the following lemma.

**Lemma 7.** *Let  $\varepsilon$  be any positive constant. Then, given a set of votes  $\mathcal{V}$ , there exists a subset  $\mathcal{V}' \subset \mathcal{V}$  of size  $O\left(\frac{\log m}{\varepsilon^2}\right)$  such that  $h(\mathcal{V}, c(\mathcal{V}')) \leq (1 + \varepsilon)h(\mathcal{V}, c(\mathcal{V}))$ , where  $c(\mathcal{V})$  and  $c(\mathcal{V}')$  are the committees chosen by Minisum voting rule on  $\mathcal{V}$  and  $\mathcal{V}'$  respectively.*

*Proof.* We prove the existence of a subset  $\mathcal{V}' \subset \mathcal{V}$  of size  $O(\log m/\varepsilon^2)$  such that  $h(\mathcal{V}, c(\mathcal{V}')) \leq (1 + 20\varepsilon)h(\mathcal{V}, c(\mathcal{V}))$ . Applying the weaker inequality with  $\varepsilon' = \frac{\varepsilon}{20}$  then proves the Lemma. We assume without loss of generality that  $\varepsilon \leq \frac{1}{40}$ . We prove the statement using probabilistic methods. We pick  $r$  votes from  $\mathcal{V}$  uniformly at random to form the set  $\mathcal{V}'$ . Let  $c \in \mathcal{C}$  be any arbitrary candidate. Let  $\mathbb{1}_{\mathcal{V}}(c)$  and  $\mathbb{1}_{\mathcal{V}'}(c)$  be the number of approvals that the candidate  $c$  receives from the votes in  $\mathcal{V}$  and  $\mathcal{V}'$  respectively. There exist an  $r = O(\log m/\varepsilon^2)$  such that the following holds (Theorem 5 in [Dey and Bhattacharyya, 2015]).

$$\Pr \left[ \forall c \in \mathcal{C}, \left| \frac{n}{r} \mathbb{1}_{\mathcal{V}'}(c) - \mathbb{1}_{\mathcal{V}}(c) \right| \leq \varepsilon n \right] > 0$$

Now assume that for every candidate  $c \in \mathcal{C}$ , we have  $\left| \frac{n}{r} \mathbb{1}_{\mathcal{V}'}(c) - \mathbb{1}_{\mathcal{V}}(c) \right| \leq \varepsilon n$  and the statement above shows that there exists such a set  $\mathcal{V}'$ . With this assumption, we now prove the result. Suppose  $c(\mathcal{V}') \setminus c(\mathcal{V}) = \{z_i : i \in [\ell]\}$ . Then there exists exactly  $\ell$  many candidates  $z'_i, i \in [\ell]$  such that,  $c(\mathcal{V}) \setminus c(\mathcal{V}') = \{z'_i : i \in [\ell]\}$  since both  $c(\mathcal{V})$  and  $c(\mathcal{V}')$  are committees of same size  $m^*$ . The assumption that for every candidate  $c \in \mathcal{C}$ , we have  $\left| \frac{n}{r} \mathbb{1}_{\mathcal{V}'}(c) - \mathbb{1}_{\mathcal{V}}(c) \right| \leq \varepsilon n$  gives us the following:

$$h(\mathcal{V}, c(\mathcal{V}')) \leq h(\mathcal{V}, c(\mathcal{V})) + 4\varepsilon\ell n$$

Now we consider the following two cases.

- Case 1:  $\mathbb{1}(z'_i) \leq \frac{n}{4}$  for every  $i \in [\ell]$  :  
We have  $h(\mathcal{V}, c(\mathcal{V})) \geq \frac{3n}{4}\ell$  and thus  $h(\mathcal{V}, c(\mathcal{V}')) \leq (1 + \frac{16}{9}\varepsilon)h(\mathcal{V}, c(\mathcal{V}))$ .
- Case 2:  $\mathbb{1}(z'_i) \geq \frac{n}{4}$  for some  $i \in [\ell]$  :

For every  $j \in [\ell]$ , the candidate  $z_j$  is in the committee  $c(\mathcal{V}')$  implies the following.

$$\begin{aligned}
\mathbb{1}(z_i) &\geq \frac{n}{4} - 2\varepsilon n \text{ for every } i \in [\ell] \\
\Rightarrow h(\mathcal{V}, c(\mathcal{V})) &\geq \left(\frac{n}{4} - 2\varepsilon n\right)\ell \\
&\geq \frac{n}{5}\ell \\
\Rightarrow h(\mathcal{V}, c(\mathcal{V}')) &\leq (1 + 20\varepsilon)h(\mathcal{V}, c(\mathcal{V}))
\end{aligned}$$

The third line follows from the assumption that  $\varepsilon \leq \frac{1}{40}$ .

□

We now turn to our approximation algorithm.

**Theorem 6.** *There is a  $(1 + \varepsilon)$ -approximation algorithm for the MINISUM-OUTLIERS problem running in time  $n^{O(\log m/\varepsilon^2)}$ , for every constant  $\varepsilon > 0$ .*

*Proof.* Let  $\mathcal{V}^* \subset \mathcal{V}$  be the set of non-outliers in the optimal solution. We apply Lemma 7 to the set  $\mathcal{V}^*$ . The algorithm guesses the set  $\mathcal{V}'$  by trying all possible  $n^{O(\log m/\varepsilon^2)}$  subsets of  $\mathcal{V}$  (since we do not know  $\mathcal{V}^*$ ). Let  $c(\mathcal{V}')$  be the Minisum committee of the votes in  $\mathcal{V}'$ . The algorithm returns the set  $\mathcal{V}''$  of the  $n^*$  votes that have the smallest Hamming distances from  $c(\mathcal{V}')$  and the committee  $c(\mathcal{V}'')$ . Then we have the following.

$$h(c(\mathcal{V}''), \mathcal{V}'') \leq h(c(\mathcal{V}'), \mathcal{V}'') \leq h(c(\mathcal{V}'), \mathcal{V}^*) \leq (1 + \varepsilon)h(c(\mathcal{V}^*), \mathcal{V}^*)$$

The last inequality follows from Lemma 7.

□

## 5.2 Hardness of Approximation

In the previous section, we demonstrated some approaches for approximating the target score in the context of minisum voting rule. We now show that the other voting rules are inapproximable in a very strong sense. Our first hardness result is that it is NP-hard to check even if there is a subset of outliers and a corresponding committee for which the net disapproval score is zero. This follows from Lemma 6.

**Corollary 2.** *There does not exist any polynomial time  $\alpha(m, n)$ -approximation algorithm for the NET-DISAPPROVAL-OUTLIERS problem for any computable function  $\alpha(\cdot, \cdot)$ , unless  $P = NP$ .*

Next, we show that it is NP-hard to check even if there is a subset of outliers and a corresponding committee for which the disapproval score is zero. For this, we reduce from a problem closely related to MAX-CLIQUE, namely that of finding a largest-sized biclique in a graph  $G$ , which is known to be NP-complete [Garey and Johnson, 1979, Johnson, 1987].

**BICLIQUE**

**Input:** An undirected bipartite graph  $\mathcal{G} = (V_L, V_R, E)$  and an integer  $k$

**Question:** Do there exist two subsets of vertices  $A \subset V_L$  and  $B \subset V_R$  with  $|A| = |B| = k$  such that  $(A, B)$  forms a biclique?

**Lemma 8.** *There does not exist any polynomial time  $\alpha(m, n)$ -approximation algorithm for DISAPPROVAL-OUTLIERS for any computable function  $\alpha(\cdot, \cdot)$ , unless  $P = NP$ .*

*Proof.* We will show that deciding whether the optimal disapproval score is zero or not, is NP-complete, thereby implying the result. The problem is clearly in NP. To prove NP-hardness, we reduce the BICLIQUE problem to the DISAPPROVAL-OUTLIERS problem. Let  $(\mathcal{G} = (V_L, V_R, E), k)$  be an arbitrary instance of the BICLIQUE problem. Let  $V_L = \{u_1, \dots, u_n\}$  and  $V_R = \{u'_1, \dots, u'_m\}$ . We define the corresponding instance  $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$  of the DISAPPROVAL-OUTLIERS problem as follows:

$$\begin{aligned}\mathcal{V} &= \{\mathcal{S}_1, \dots, \mathcal{S}_n\}, \mathcal{C} = \{c_1, \dots, c_m\}, \\ \mathcal{S}_i &= \{c_j : (u_i, u'_j) \in E\} \quad \forall i \in [n], n^* = k, m^* = k, t = 0\end{aligned}$$

We claim that the two instances are equivalent. In the forward direction, suppose  $A \subset V_L$  and  $B \subset V_R$  with  $|A| = |B| = k$  forms a biclique. Consider the set of non-outliers to be  $\mathcal{V}^* = \{\mathcal{S}_i : u_i \in A\}$  and the committee  $\mathcal{C}^* = \{c_j : u'_j \in B\}$ . This achieves disapproval score of 0.

In the reverse direction, suppose there exist a set of non-outliers  $\mathcal{V}^*$  and a committee  $\mathcal{C}^*$  which achieves disapproval score of 0. Without loss of generality, we assume that  $|\mathcal{V}^*| = k$  since removing votes from the set of non-outliers can only reduce the score. Consider  $A = \{u_i : \mathcal{S}_i \in \mathcal{V}^*\} \subset V_L$  and  $B = \{c_j : c_j \in \mathcal{C}^*\} \subset V_R$ . We claim that  $A$  and  $B$  must form a biclique. Indeed, otherwise the disapproval score would be strictly more than 0.  $\square$

## 6 Concluding Remarks and Future Directions

We have studied the problem of determining the a committee that is good for a large portion of the electorate, based on some common voting rules. We begin this section by showing that our choice of the notion of disapproval was an arbitrary one, by demonstrating the following lemma. We use the notation  $\text{disapproval}'$  to denote the voting rule given by the corresponding bottom row in Table 1.

**Lemma 9.** *There is a polynomial time approximation preserving reduction from DISAPPROVAL-OUTLIERS problem to DISAPPROVAL'-OUTLIERS problem and vice-versa. Moreover, the reductions are such that  $m^*$  in the original instance becomes  $\overline{m}$  in the reduced instance,  $\overline{m}$  in the original instance becomes  $m^*$  in the reduced instance, and every other parameter remain exactly same.*

*Proof.* We provide the reduction from DISAPPROVAL-OUTLIERS to DISAPPROVAL'-OUTLIERS. The other cases are exactly same. Let  $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$  be an arbitrary instance of DISAPPROVAL-OUTLIERS. We define the corresponding instance  $(\mathcal{V}', \mathcal{C}', n'^*, m'^*, t')$  of DISAPPROVAL'-OUTLIERS as follows.

$$\mathcal{C}' = \mathcal{C}, \mathcal{S}'_i = \{x \in \mathcal{C}' : x \notin \mathcal{S}_i\}, n'^* = n^*, m'^* = \overline{m}, t' = t$$

We claim that the two instances are equivalent. If there exist a  $\mathcal{V}^* \subset \mathcal{V}$  and  $\mathcal{C}^* \subset \mathcal{C}$  such that  $s_{\text{disapproval}}(\mathcal{V}^*, \mathcal{C}^*) \leq t$  then  $s_{\text{disapproval}'}(\mathcal{V}^*, \overline{\mathcal{C}}) \leq t$  and vice versa.  $\square$

There are some very special cases that are solvable in polynomial time. In particular, when every candidate is approved by at most one vote, or when every vote approves at most one candidate, then the MINISUM-OUTLIERS problem can be solved in polynomial amount of time.

**Lemma 10.** *The MINISUM-OUTLIERS problem is in P when every vote approves at most one candidate.*

*Proof.* The following greedy algorithm computes the committee and the set of outliers which achieves lowest minisum score. First choose the  $m^*$  candidates with maximum number of approvals. Let  $\mathcal{C}^*$  be the chosen set of  $m^*$  candidates. Let  $U$  be the set of votes who approve some candidate in  $\mathcal{C}^*$ ,  $W$  be the set of votes who approve some candidate in  $\overline{\mathcal{C}} (= \mathcal{C} \setminus \mathcal{C}^*)$ , and  $Y = \mathcal{V} \setminus (U \cup W)$ . Then, if  $|W| \geq \overline{n}$ , then define the set of outliers to be  $\overline{\mathcal{V}}$  for any arbitrary  $\overline{\mathcal{V}} \subseteq W$  with  $|\overline{\mathcal{V}}| = \overline{n}$ . If  $|W| < \overline{n}$  and  $|W| + |Y| \geq \overline{n}$ , then define the set of outliers to be  $\overline{\mathcal{V}} = W \cup Z$ , for any arbitrary  $Z \subseteq Y$  with  $|Z| = \overline{n} - |W|$ . Otherwise, define the set of outliers to be  $W \cup Y \cup A$ , for any arbitrary  $A \subseteq U$  with  $|A| = \overline{n} - |W| + |Y|$ .  $\square$

**Lemma 11.** *The MINISUM-OUTLIERS problem is in P when every candidate is approved in at most one vote*

*Proof.* Consider the following dynamic programming based algorithm. Let the votes be  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_n$ . We define  $M(i, j, \ell)$  to be a set of  $j$  outliers (call it an optimal set of outliers) that minimizes the sum of Hamming distances when we are selecting a committee of size  $\ell$  considering only votes in  $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_i\}$ . We compute the value of  $M(i, j, \ell)$ . Consider the case when the vote  $\mathcal{S}_i$  belongs to an optimal set of outliers when we are selecting a committee of size  $\ell$  considering only votes in  $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_i\}$ . In this case,  $M(i, j, \ell) = M(i - 1, j - 1, \ell)$ . Consider the other case when the vote  $\mathcal{S}_i$  does not belong to any optimal set of outliers when we are selecting a committee of size  $\ell$  considering only votes in  $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_i\}$ . Then  $M(i, j, \ell) = M(i - 1, j, \ell - |\mathcal{S}_i|)$  if  $\ell \geq |\mathcal{S}_i|$  and  $M(i, j, \ell) = M(i - 1, j, 0)$  otherwise, since when the vote  $\mathcal{S}_i$  is not an outlier then we can assume without loss of generality that the committee has maximum overlap with the vote  $\mathcal{S}_i$ . Base cases are taken care from the fact that  $M(i, j, \ell)$  can be computed in polynomial time irrespective of the values of  $\ell$  and  $j$  when  $i$  is a constant. We compute the value of  $M(i, j, \ell)$  for every  $i \in [n], j \leq i, \ell \in [m]$ .  $M(n, \bar{n}, m^*)$  gives us the optimal set of outliers (which also in turn gives us the committee to choose).  $\square$

**Future Directions** One pertinent open problem is to close the only unresolved case of parameterized complexity of  $s$ -OUTLIERS for the disapproval voting rule when parameterized by  $(\overline{m}, \overline{n}, t)$ . It is also open to improve, or prove lower bounds, for our approximation results in the context of the minsum voting rule. Refining the FPT fragment of our dichotomy to kernelization is also an exciting direction for further investigation. Finally, we are also interested in extending our dichotomous results to a dichotomy based on more parameters. We have made progress towards that by incorporating two more parameters, namely the maximum number of candidates that any vote approves and maximum number of approvals that any candidate obtains.



## References

- [Abrahamson et al., 1995] Abrahamson, K. R., Downey, R. G., and Fellows, M. R. (1995). Fixed-parameter tractability and completeness IV: on completeness for  $W[P]$  and PSPACE analogues. *Annals of Pure and Applied Logic*, 73(3):235–276.
- [Aziz et al., 2015] Aziz, H., Brill, M., Conitzer, V., Elkind, E., Freeman, R., and Walsh, T. (2015). Justified representation in approval-based committee voting. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 784–790.
- [Bartholdi et al., 1992] Bartholdi, J. J., Tovey, C. A., and Trick, M. A. (1992). How hard is it to control an election? *Mathematical and Computer Modelling*, 16(8):27–40.
- [Bartholdi III et al., 1989a] Bartholdi III, J., Tovey, C. A., and Trick, M. A. (1989a). Voting schemes for which it can be difficult to tell who won the election. *Social Choice and welfare*, 6(2):157–165.
- [Bartholdi III et al., 1989b] Bartholdi III, J. J., Tovey, C. A., and Trick, M. A. (1989b). The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241.
- [Baumeister et al., 2010] Baumeister, D., Erdélyi, G., Hemaspaandra, E., Hemaspaandra, L. A., and Rothe, J. (2010). Computational aspects of approval voting. In *Handbook on Approval Voting*, pages 199–251. Springer.
- [Betzler et al., 2010] Betzler, N., Guo, J., and Niedermeier, R. (2010). Parameterized computational complexity of dodgson and young elections. *Information and Computation*, 208(2):165–177.
- [Betzler and Uhlmann, 2008] Betzler, N. and Uhlmann, J. (2008). Parameterized complexity of candidate control in elections and related digraph problems. In *Combinatorial Optimization and Applications*, pages 43–53. Springer.
- [Boucher et al., 2013] Boucher, C., Landau, G. M., Levy, A., Pritchard, D., and Weimann, O. (2013). On approximating string selection problems with outliers. *Theoretical Computer Science*, 498:107–114.
- [Boucher and Ma, 2011] Boucher, C. and Ma, B. (2011). Closest string with outliers. *BMC Bioinformatics*, 12(S-1):S55.
- [Brams, 2008] Brams, S. J. (2008). Mathematics and democracy: Designing better voting and fair-division procedures. *Mathematical and Computer Modelling*, 48(9-10):1666–1670.
- [Brams et al., 2007] Brams, S. J., Kilgour, D. M., and Sanver, M. R. (2007). A minimax procedure for electing committees. *Public Choice*, 132(3-4):401–420.
- [Brandt et al., 2012] Brandt, F., Conitzer, V., and Endriss, U. (2012). Computational social choice. *Multi-agent systems*, pages 213–283.
- [Byrka and Sornat, 2014] Byrka, J. and Sornat, K. (2014). PTAS for minimax approval voting. In *Web and Internet Economics*, volume 8877 of *Lecture Notes in Computer Science*, pages 203–217. Springer.
- [Caragiannis et al., 2010] Caragiannis, I., Kalaitzis, D., and Markakis, E. (2010). Approximation algorithms and mechanism design for minimax approval voting. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*.
- [Chevaire et al., 2007] Chevaire, Y., Endriss, U., Lang, J., and Maudet, N. (2007). *A short introduction to computational social choice*. Springer.
- [Conitzer, 2010] Conitzer, V. (2010). Making decisions based on the preferences of multiple agents. *Commun. ACM*, 53(3):84–94.

- [Conitzer et al., 2007] Conitzer, V., Sandholm, T., and Lang, J. (2007). When are elections with few candidates hard to manipulate? *Journal of the ACM (JACM)*, 54(3):14.
- [Dey and Bhattacharyya, 2015] Dey, P. and Bhattacharyya, A. (2015). Sample complexity for winner prediction in elections. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015, Istanbul, Turkey, May 4-8, 2015*, pages 1421–1430.
- [Downey and Fellows, 1995a] Downey and Fellows (1995a). Fixed-parameter tractability and completeness I: Basic results. *SICOMP: SIAM Journal on Computing*, 24.
- [Downey and Fellows, 1995b] Downey and Fellows (1995b). Fixed-parameter tractability and completeness II: On completeness for  $W[1]$ . *TCS: Theoretical Computer Science*, 141.
- [Downey and Fellows, 1988] Downey, R. and Fellows, M. (1988). Fixed-parameter tractability and completeness III: Some structural aspects of the  $W$  hierarchy. In Ambos-Spies, K., Homer, S., and Schöning, U., editors, *Complexity Theory: Current Research*, pages 191–225, Cambridge. Cambridge University Press.
- [Downey and Fellows, 1999a] Downey, R. G. and Fellows, M. R. (1999a). *Parameterized complexity*. Springer-Verlag, New York.
- [Downey and Fellows, 1999b] Downey, R. G. and Fellows, M. R. (1999b). *Parameterized Complexity*, volume 3. springer Heidelberg.
- [Downey and Fellows, 2013] Downey, R. G. and Fellows, M. R. (2013). *Fundamentals of Parameterized Complexity*. Springer Publishing Company, Incorporated.
- [Elkind et al., 2009] Elkind, E., Faliszewski, P., and Slinko, A. (2009). On distance rationalizability of some voting rules. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 108–117. ACM.
- [Faliszewski et al., 2011] Faliszewski, P., Hemaspaandra, E., and Hemaspaandra, L. A. (2011). Multimode control attacks on elections. *Journal of Artificial Intelligence Research*, 40(1):305–351.
- [Faliszewski et al., 2009] Faliszewski, P., Hemaspaandra, E., Hemaspaandra, L. A., and Rothe, J. (2009). A richer understanding of the complexity of election systems. In *Fundamental Problems in Computing*, pages 375–406. Springer.
- [Fishburn, 1978] Fishburn, P. C. (1978). Axioms for approval voting: Direct proof. *Journal of Economic Theory*, 19(1):180–185.
- [Garey and Johnson, 1979] Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability*, volume 174. freeman New York.
- [Gramm et al., 2003] Gramm, J., Niedermeier, R., Rossmanith, P., et al. (2003). Fixed-parameter algorithms for closest string and related problems. *Algorithmica*, 37(1):25–42.
- [Hemaspaandra et al., 2007] Hemaspaandra, E., Hemaspaandra, L. A., and Rothe, J. (2007). Anyone but him: The complexity of precluding an alternative. *Artificial Intelligence*, 171(5):255–285.
- [Johnson, 1987] Johnson, D. S. (1987). The np-completeness column: An ongoing guide. *J. Algorithms*, 8(5):438–448.
- [Kilgour, 2010] Kilgour, D. M. (2010). Approval balloting for multi-winner elections. In *Handbook on approval voting*, pages 105–124. Springer.
- [LeGrand et al., 2007] LeGrand, R., Markakis, E., and Mehta, A. (2007). Some results on approximating the minimax solution in approval voting. In *AAMAS*, volume 7, pages 1185–1187.

- [Lepelley, 2013] Lepelley, D. (2013). Jean-françois laslier and M. remzi sanver (eds.): Handbook on approval voting. *Social Choice and Welfare*, 40(4):1231–1235.
- [Lo et al., 2014] Lo, C., Kakaradov, B., Lokshtanov, D., and Boucher, C. (2014). SeeSite: characterizing relationships between splice junctions and splicing enhancers. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(4):648–656.
- [Meir et al., 2008] Meir, R., Procaccia, A. D., Rosenschein, J. S., and Zohar, A. (2008). Complexity of strategic behavior in multi-winner elections. *J. Artif. Intell. Res. (JAIR)*, 33:149–178.
- [Misra et al., 2015] Misra, N., Nabeel, A., and Singh, H. (2015). On the parameterized complexity of min-imax approval voting. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015, Istanbul, Turkey, May 4-8, 2015*, pages 97–105.
- [Procaccia et al., 2007] Procaccia, A. D., Rosenschein, J. S., and Zohar, A. (2007). Multi-winner elections: Complexity of manipulation, control and winner-determination. In *IJCAI*, volume 7, pages 1476–1481.
- [Procaccia et al., 2008] Procaccia, A. D., Rosenschein, J. S., and Zohar, A. (2008). On the complexity of achieving proportional representation. *Social Choice and Welfare*, 30(3):353–362.
- [Rothe et al., 2003] Rothe, J., Spakowski, H., and Vogel, J. (2003). Exact complexity of the winner problem for young elections. *Theory of Computing Systems*, 36(4):375–386.
- [Skowron and Faliszewski, 2015] Skowron, P. K. and Faliszewski, P. (2015). Fully proportional representation with approval ballots: Approximating the maxcover problem with bounded frequencies in FPT time. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 2124–2130.
- [Young and Levenglick, 1978] Young, H. P. and Levenglick, A. (1978). A consistent extension of condorcet’s election principle. *SIAM Journal on applied Mathematics*, 35(2):285–300.

## Appendix

Now we provide the missing proofs.

**Proposition 1.** *Let  $s \in \{\text{Minisum}, \text{Disapproval}, \text{Net-Disapproval}\}$ . Then, there is a  $O^*(2^m)$  time algorithm and a  $O^*(2^n)$  time algorithm for the  $s$ -OUTLIERS problem.*

*Proof.* Notice that if we know the committee  $\mathcal{C}^* \subset \mathcal{C}$  in the optimal solution then we can solve the problem in polynomial amount of time by choosing  $n^*$  many votes as non-outliers that have the smallest score with respect to the committee  $\mathcal{C}^*$ . We guess the committee  $\mathcal{C}^*$  of the optimal solution (by iterating over  $O(2^m)$  many possibilities) and solve the problem in time  $O^*(2^m)$ .

Notice that if we know the set of non-outliers  $\mathcal{V}^* \subset \mathcal{V}$  in the optimal solution then we can solve the problem in polynomial amount of time by choosing  $m^*$  many candidates that receive least disapprovals from the votes in  $\mathcal{V}^*$ . We guess the set of non-outliers  $\mathcal{V}^*$  of the optimal solution (by iterating over  $O(2^n)$  many possibilities) and solve the problem in time  $O^*(2^n)$ .  $\square$

**Lemma 3.** *Let  $s \in \{\text{Minisum}, \text{Net-Disapproval}, \text{Disapproval}\}$ . Then the  $s$ -OUTLIERS problem is  $W[1]$ -hard, when parameterized by  $(t + m^* + n^*)$ , even when every vote approves exactly two candidates.*

*Proof.* First let us prove the result for the MINISUM-OUTLIERS problem. We exhibit a parameterized reduction from the CLIQUE problem to the MINISUM-OUTLIERS problem thereby proving the result. Let  $(\mathcal{G} =$

$(U, E), k$  be an arbitrary instance of the CLIQUE problem. Let  $U = \{u_1, \dots, u_n\}$  and  $E = \{e_1, \dots, e_m\}$ . We define the corresponding instance  $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$  of the MINISUM-OUTLIERS problem as follows:

$$\mathcal{V} = \{\mathcal{S}_1, \dots, \mathcal{S}_m\}, \mathcal{C} = \{c_1, \dots, c_n\}, n^* = \binom{k}{2}, m^* = k,$$

$$t = (k-2) \binom{k}{2}, \mathcal{S}_i = \{c_j : v_j \in e_i\}, \forall i \in [m]$$

We claim that the two instances are equivalent. In the forward direction, suppose  $W \subset U$  forms a clique with  $|W| = k$  in  $\mathcal{G}$ , and let  $Q$  denote the set of edges that have both endpoints in  $W$ . Consider the committee  $\mathcal{C}^* = \{c_i : u_i \in W\}$  and the set of non-outliers  $\mathcal{V}^* = \{\mathcal{S}_i : e_i \in Q\}$ . This achieves the minisum score of  $(k-2) \binom{k}{2}$ .

In the reverse direction, suppose there exist a set of non-outliers  $\mathcal{V}^* \subset \mathcal{V}$  with  $|\mathcal{V}^*| \geq \binom{k}{2}$  and a committee  $\mathcal{C}^* \subset \mathcal{C}$  with  $|\mathcal{C}^*| = k$  such that  $h(\mathcal{V}^*, \mathcal{C}^*) \leq (k-2) \binom{k}{2}$ . Without loss of generality, we assume  $|\mathcal{V}^*| = \binom{k}{2}$  since more votes in the set of non-outliers cannot reduce the minisum score. We claim that all the edges in  $Q = \{e_j : \mathcal{S}_j \in \mathcal{V}^*\}$  have both the end points in  $W = \{u_i : c_i \in \mathcal{C}^*\}$  thereby the vertices in  $W$  forming a clique. Indeed otherwise, suppose that there is an edge  $e_j \in Q$  for which both the end points are not in  $W$ . Then  $h(\mathcal{S}_j, \mathcal{C}^*) > (k-2)$ . On the other hand we have  $h(\mathcal{S}, \mathcal{C}^*) \geq (k-2)$  for every  $\mathcal{S} \in \mathcal{V}^*$ . Hence we have  $h(\mathcal{V}^*, \mathcal{C}^*) > (k-2) \binom{k}{2}$  which is a contradiction.

The proofs for the other rules are identical, except for the values of the target score. We define  $t = (k-2) \binom{k}{2}$  for the disapproval and  $t = (k-4) \binom{k}{2}$  for the net disapproval voting rules. It is easily checked that the details are analogous.  $\square$

**Theorem 5.** *There is a  $\varepsilon \overline{m}$ -approximation algorithm for the MINISUM-OUTLIERS problem, for every constant  $\varepsilon > 0$ .*

*Proof.* Fix any constant  $\varepsilon > 0$ . Consider the case that the minisum committee, say  $\mathcal{C}^*$ , incurs a sum of Hamming distances of at most  $\frac{n^*}{\varepsilon}$ . In this case, by pigeon hole principle, there exists a vote say  $\mathcal{S}_i$  whose Hamming distance from the committee  $\mathcal{C}^*$  is at most  $\frac{1}{\varepsilon}$ . We can guess the vote  $\mathcal{S}_i$  (by iterating over all the votes) and try all possible committees of size  $m^*$  which is at most  $\frac{1}{\varepsilon}$  far away from  $\mathcal{S}_i$ ; this takes polynomial amount of time since there are only  $O(m^{\frac{1}{\varepsilon}})$  many such committees. Hence, in polynomial amount of time, we can check if the sum of Hamming distances of the minisum committee and the votes is less than  $\frac{n^*}{\varepsilon}$  or not. Now assume that  $\mathcal{C}^*$  incurs a sum of the Hamming distances of at least  $\frac{n^*}{\varepsilon}$ . We choose any  $n^*$  many voters and return the committee  $\mathcal{X}$  consisting of  $m^*$  many candidates that receive most approvals. Notice that the sum of the number of disapprovals that any candidate in the committee receives and the number of approvals that any candidate not in the committee receives is at most  $n^*$ . Hence, the sum of Hamming distances of  $\mathcal{X}$  and the  $n^*$  many votes chosen is at most  $m^* n^* + (m - 2m^*) n^* = \overline{m} n^*$ .  $\square$

**Lemma 11.** *The MINISUM-OUTLIERS problem is in P when every candidate is approved in at most one vote*

*Proof.* Consider the following dynamic programming based algorithm. Let the votes be  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_n$ . We define  $M(i, j, \ell)$  to be a set of  $j$  outliers (call it an optimal set of outliers) that minimizes the sum of Hamming distances when we are selecting a committee of size  $\ell$  considering only votes in  $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_i\}$ . We compute the value of  $M(i, j, \ell)$ . Consider the case when the vote  $\mathcal{S}_i$  belongs to an optimal set of outliers when we are selecting a committee of size  $\ell$  considering only votes in  $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_i\}$ . In this case,  $M(i, j, \ell) = M(i-1, j-1, \ell)$ . Consider the other case when the vote  $\mathcal{S}_i$  does not belong to any optimal set of outliers when we are selecting a committee of size  $\ell$  considering only votes in  $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_i\}$ . Then  $M(i, j, \ell) = M(i-1, j, \ell - |\mathcal{S}_i|)$  if  $\ell \geq |\mathcal{S}_i|$  and  $M(i, j, \ell) = M(i-1, j, 0)$  otherwise, since when the vote  $\mathcal{S}_i$  is not an outlier then we can assume without loss of generality that the committee has maximum overlap with the vote  $\mathcal{S}_i$ . Base cases are taken care from the fact that  $M(i, j, \ell)$  can be computed in polynomial

time irrespective of the values of  $\ell$  and  $j$  when  $i$  is a constant. We compute the value of  $M(i, j, \ell)$  for every  $i \in [n], j \leq i, \ell \in [m]$ .  $M(n, \bar{n}, m^*)$  gives us the optimal set of outliers (which also in turn gives us the committee to choose).  $\square$